Becoming a Data Scientist

Welcome back to the last part of our data science course,
before we do our demo.
So this piece is gonna be just about becoming a data scientist.
Hopefully at this point,
you're like that sounds like it's a pretty cool field to be in.
And I absolutely agree.
So, one of the cool things about data science,
even though it's been a growing field for the last few years,
it's still a small enough field that it feels like a community.
And you can really become part of that community.
It's very welcoming, people are out there trying to figure out best
ways to deal with data, and looking for new information, right?
I mean, everybody in this field is interested in data.
And more data points in general are better.
And so, you can be one of those data points helping out the community.
And so, it's a great place to be, great time to do it.
So, what is the path to becoming a data scientist?
So, there's kind of this classical path.
It's like okay, so we're gonna learn statistics.
We're gonna get really good at statistics.
Become professional statisticians.
Then, we're gonna take some computer programming and
apply some computer programming to that.
We're gonna then jump into machine learning, and
then we're gonna practice.
And I think that's a really easy path to follow,
but I'm not sure it's the best path to follow.
I think there's this really interesting thing that I've seen as
an educator, that when you jump into something and kind of follow
this classical path, and take course after course after course.
Depending on good of a student you are, you definitely learn
this stuff, but you don't have any application to it.
And so, my view of becoming a data scientist,
is as follows, if you're totally new to the field.
Definitely, you wanna learn these core pieces.
So you wanna learn some statistics,
you wanna learn some computer programming.
But you don't have to be masters at them.
What you wanna do, is you wanna find some data to love and explore.
So, you want to kinda understand these core basics.
Kinda understand level one stuff.
Maybe even dabble in some of the level three machine learning stuff,
and neural nets and whatnot.
But then go and say I'm gonna attack this data set.

And attack it, work with the community.
You can easily go out there and talk to colleagues,
send out stuff on Stack Overflow, talk about what you're trying to do,
go to places like Taggle and enter competitions.
And join the community, and say I'm attacking this thing In x, y, or z.
I'm using k-means clustering,
I'm using support vector machines, trying to do this thing.
I don't understand this piece of it, or be open to that learning.
And I think that's a really good way to kind of
really accelerate your learning.
And then once you kind of have that advanced core of knowledge,
then go back and start taking the classes that you really wanna take.
So, my view of it is, and learning in general,
I think you can learn so much more if you're passionate about it,
than you do if you're just kind of going through the motions.
And, so don't just go through the motions.
Don't just take class, after class,
after class if you don't really love it.
And one way to love it, is find something to do with it right away.
And so, my advice to you is,
as I said, learn the basics, find some data and just go for it.
But if you like the more academic structure,
there are a number of programs out there.
Again, I recommend the Microsoft Virtual Academy course
that's specifically on data science.
But there's wonderful YouTube videos out there about all sorts of things,
from machine learning to statistics that you can go through.
So if you wanted to look at some data sets, UCI,
University of California Irvine, has some great data sets.
Check out their stuff.
And that's what we're gonna be dealing with when we do our
demo on Excel.
And with that, I think I just wanna go through a few takeaways.
So, data science is really a scientific process.
We're really applying the scientific method to things.
We're taking data, we're asking questions about that data,
we're making hypotheses about that data, we're running experiments,
we're analyzing it, we're getting results out of that data.
And, there's kind of the, as I said, kind of
the white tower side of things, and the applied side of things.
But both of them really take an analytical approach to how we're
approaching our data.
So, hopefully you kind of have a general sense of what data science
is, after this course.
Then, I think the next kind of step is,
you really want to start gaining data intuition.

And so, figuring out what does a Gaussian curve actually mean?
What is a t-test?
All of these things I think, are really kind of the next step.
Then, one thing I think a lot of us tend to do when we're learning
something new, is we try to take everything at once.
And it's good to kind of get that big overview, but
really dive in on a piece, and learn it incrementally.
So, take a piece and spend an hour, really sitting down and just,
I'm gonna learn all I can about, coefficient of determination.
And just stay focused on it.
So, that's another piece that I wanted you to take away from this.
And the last piece is, to keep that focus and keep that drive,
you really need to be passionate about what you're trying to learn.
And so, if you don't have that passion,
I think it's really hard to learn.
Even if you're like I really wanna do these amazing things as
a data scientist, cuz they have great jobs.
That's not really the right way to become a data scientist.
It's I really wanna deal with this data,
because it's something that I'm excited about,
rather than the end point of getting a great job or whatever, so.
With that, I think what we're gonna move on to is a demo, just kind of
talking about some of the level one tools that you can do in Excel.
Again, this is a piece that some of you may know already.
If you do, it's okay to move on and
jump in and start learning about data science on your own.
If you're at the point where you kinda just wanna see how somebody
might attack it, I think this next chapter will be pretty fun to watch.
So, thanks again.